3Б

УДК 004.934.2

Е.А. Бондаренко, О.А. Каплина
Институт проблем искусственного интеллекта
МОН Украины и НАН Украины, г. Донецк
Украина, 83048, г. Донецк, ул. Артема, 118 б

Принципы автоматической обработки естественно-языковых текстов: валентностный подход

E.A. Bondarenko, O.A. Kaplina
Institute of Artificial Intelligence
MES of Ukraine and NAS of Ukraine, c. Donetsk
Ukraine, 83048, c. Donetsk, Artema st., 118 b

Principles of Automatic Processing of Natural Language Texts: Valentnostny Approach

О.О. Бондаренко, О.О. Капліна
Інститут проблем штучного інтелекту
МОН України і НАН України, м. Донецьк
Україна, 83048, м. Донецьк, вул. Артема 118 б

Принципи автоматичної обробки природно-мовних текстів: валентнісний підхід

В данной работе были разработаны основные правила валентностной реализации глагольных предикатов определённых семантических классов в предложениях русского языка. С этой целью были проанализированы принципы прогнозирования количества и качества актантов, особенности их формальной и семантической реализации в естественно-языковом тексте и разработана система минимальных структурных схем предложений.
Ключевые слова: валентность, предикат, актант, естественно-языковой текст, семантическая информация.

In offered work the basic rules of valentnostny realization of verbal predicates of certain semantic classes in Russian offers were developed. The principles of forecasting of quantity and quality of actants, features of their formal and semantic realization in the natural language text were for this purpose analysed, the system of the minimum block diagrams of offers is developed.
Key Words: valency, predicate, actant, natural language text, semantic information.

У запропонованій роботі було розроблено основні правила валентнісної реалізації дієслівних предикатів певних семантичних класів у реченнях російської мови. З цією метою було проаналізовано принципи прогнозування кількості та якості актантів, особливості їхньої формальної та семантичної реалізації в природно-мовному тексті та розроблено систему мінімальних структурних схем речень.
Ключові слова: валентність, предикат, актант, природно-мовний текст, семантична інформація.

Одним из наиболее актуальных направлений в области машинной обработки естественного языка в настоящее время является разработка методов комплексного семантико-синтаксического анализа естественно-языковых текстов, то есть создание системы, наиболее полно соответствующей человеческому восприятию. Сложность

формализации неструктурированной, естественно-языковой текстовой информации состоит в представлении текста не просто в виде последовательности лексических единиц, обладающих определённой морфологической информацией, а как единого смыслового целого.

На современном этапе анализ естественно-языкового текста часто ограничивается лишь хранением, передачей, поиском и проверкой орфографии, а именно работой с формальной стороной текста. Синтагмы, полученные в результате членения языкового потока речи, анализируются чаще всего пословно, без учета семантико-синтаксических связей между членами предложения. Общая наука о построении предложения и машинное распознавание синтаксических конструкций в данном случае фактически не пересекаются, то есть информационное наполнение текста остается недоступным для большинства современных систем.

Однако стоит сказать о некоторых современных системах обработки естественно-языковых текстов, в основу которых был положен именно семантический принцип, что позволило существенно усовершенствовать обработку текстов. Одним из первых ученых, обосновавших необходимость использования семантической информации для машинной обработки текста, была проф. Н.Н. Леонтьева – специалист в области структурной и прикладной лингвистики. Её разработки стали основой прикладной машинной семантики. Позднее были созданы такие системы, как информационно-аналитическая система «Arion», позволяющая работать с разнородными источниками информации и использующая инновационные технологии обработки естественно-языковой информации [1]. Система «TextAnalyst», сходная с «Arion», разработана для анализа содержания текстов, смыслового поиска информации [2]. Названные системы, безусловно, значительно расширили возможности машинной обработки естественно-языковых текстов, однако, основываясь главным образом на семантических принципах, данные подходы несколько отстранялись от изучения семантико-синтаксических и семантико-грамматических связей между членами анализируемых текстовых структур, то есть они практически не реализуют синтаксического и точного семантического анализа текстовой информации и фактически не предполагают интеллектуальной обработки тестов, а тем более «извлечения знаний» [1].

Значительный вклад в развитие этого направления был сделан также рабочей группой АОТ, разрабатывающей программное обеспечение в области автоматической обработки текста на основе синтаксического анализа, а именно «построения синтаксических групп на одном морфологическом варианте одной клаузы»[3]. Однако в рамках данного подхода, предполагающего всесторонний анализ морфологических категорий и синтаксических связей обрабатываемого текста путем исследования его фрагментов, получаемых в результате сегментации, практически не учитывается семантическая информация, репрезентируемая определенным текстом, что отстраняет данный подход от реальной коммуникативной ситуации.

Для наиболее полного анализа естественно-языкового текста необходимо учитывать весь комплекс грамматических, синтаксических и семантических категорий. По нашему мнению, целостность такого анализа может быть обеспечена валентностным подходом, основывающимся на учете системных возможностей глагольных лексем сочетаться с определенным количеством аргументов, предусмотренных семантической структурой соответствующей лексемы.

Целью нашей работы является разработка и формализация основных правил реализации валентностных свойств глагольных предикатов для корректной машинной обработки естественно-языкового текста в соответствии с семантической информацией, репрезентируемой текстовой структурой.

Для достижения цели необходимо выполнить следующие **задачи**: исследовать основные типы предикатов и их грамматические и семантические характеристики, проанализировать способы и принципы связи предикатов и актантов в простых предложениях, схематически описать основные правила реализации валентностных свойств предикатов в тексте в соответствии с его семантикой.

Предлагаемый анализ должен включать в себя две составляющие – семантический анализ связей между членами предложения и определение грамматической зависимости между ними. Именно *валентность* предиката (т.е. члена предложения, который указывает на действие, отношение) определяет количество правосторонних и левосторонних членов синтаксической конструкции, а также форму каждого из них. Соответственно, **предметом исследования** являются только обязательные члены элементарного простого предложения. Факультативные члены синтаксической конструкции как не обусловленные валентностью предиката элементы и отсекаемые в ходе первичной сегментации в данной работе не рассматриваются.

Понятие валентности, заимствованное в середине прошлого века из химии, на сегодня является одним из наиболее востребованных в лингвистике. Возможность глагола сочетаться с другими словами и прогнозировать такую сочетаемость впервые отметил Л. Теньер в работе «Основы структурного синтаксиса» [4]. В современном языкознании существует значительное количество работ, посвященных проблеме валентности: С.Д. Кацнельсон [5], О.И. Москальская [6], М.Д. Степанова [7], Ю.Д. Апресян [8] и др. В ряде работ доказывается, что способность валентного прогнозирования присуща и таким самостоятельным частям речи, как прилагательное и существительное, а также служебным – предлогу и союзу (Московская семантическая школа). Активное развитие теория валентности получила в работах украинских лингвистов: П. П. Загнитко [9], Р. Выхованца [10] и др.

Исходя из того, что структуры естественного языка весьма разнообразны по уровню сложности, предусмотреть который при машинной обработке часто довольно сложно, в данной работе предлагается система семантического анализа простых синтаксических структур, образованных в результате сегментации естественно-языковых текстов.

Семантико-синтаксический анализ простого предложения, как минимальной единицы членения естественно-языкового текста, базируется на выделении элемента, способного быть предикатом. Носителями предикатной семантики выступают, в первую очередь, глаголы, далее – причастия, отглагольные существительные, прилагательные, существительные. Носители субстанциональной семантики (актанты) занимают по отношению к предикату определенные позиции (ячейки) – одну правостороннюю и от одной до 6 левосторонних, семантика, форма и место которых строго определены предикатом. Другие компоненты анализируемого сегмента будут факультативными.

Валентное прогнозирование способствует построению определённой обобщенной модели предложения или текста, на которую ориентируется говорящий. Ведь, по мнению Есперсена, любое предложение в языке, независимо от конкретных лексем, строится по определённой схеме [11, с. 17].

Формализация речевого потока говорящего сводит бесконечное множество создаваемых семантико-синтаксических структур к набору определённых прогнозируемых моделей, которые в схематическом представлении будут легко восприняты машиной.

Такой подход предусматривает создание электронного словаря валентности, позволяющего описать основные формально-семантические модели предложений. В основу такого словаря был положен «Экспериментальный синтаксический словарь» под. ред. Л.Г. Бабенко [12], который был расширен с 8 000 слов до 30 878 слов в

соответствии с «Грамматическим словарем русского языка» А. А. Зализняка. Каждая словарная единица была маркирована в соответствии с классификацией А. А. Зализняка [13], а также отнесена к определенному семантическому классу в соответствии с семантической классификацией глаголов Л. Г. Бабенко (табл. 1).

Таблица 1 – Словарь валентности (фрагмент)

Слово без ударения	Слово с ударением	Метка Зализняка	Семантический класс Бабенко
аблактировать	аблакт_ировать	св-нсв 2а	1.4.12
абонировать	абон_ировать	св-нсв 2а	3.2.2.2.
абонироваться	абон_ироваться	св-нсв 2а	3.2.2.4
абордировать	аборд_ировать	св-нсв 2а	3.4.1.
абортировать	аборт_ировать	св-нсв 2а	1.4.9.
абсолютизировать	абсолютиз_ировать	св-нсв 2а	1.6.9.
абсорбировать	абсорб_ировать	св-нсв 2а	1.4.10.
абстрагировать	абстраг_ировать	св-нсв 2а	1.6.3.
абстрагироваться	абстраг_ироваться	св-нсв 2а	1.6.1.
авансировать	аванс_ировать	св-нсв 2а	3.2.7.
авизировать	авиз_ировать	св-нсв 2а	1.8.8
авралить	авр_алить	нсв нп 4а	1.8.8.
автоматизировать	автоматиз_ировать	св-нсв 2а	3.4.7.
автоматизироваться	автоматиз_ироваться	св-нсв 2а	2.2.6.1.

Словарь также был дополнен разработанным нами приложением «Основные правила формальной реализации валентностных свойств глагольных предикатов» (табл. 2), где каждый семантический класс был проанализирован на формально-грамматическом уровне в соответствии с усовершенствованной теорией валентности предикатов А.П. Загнитко [9, с. 259-278], а также типологией регулярных реализаций структурных схем, предложенной В. А. Белошапковой [14, с. 363-466], что позволило значительно оптимизировать и ускорить работу по созданию формальных правил реализации предикатов в предложении, а также свести количество предлагаемых правил к минимуму.

Выделение и формализация основных правил валентностной реализации предикатов в простых предложениях предполагает следующие шаги:

1. Схематическое описание валентного гнезда (системы актантов) глагола определённого семантического класса как предиката (табл. 2 графа «*валентность*»).

2. Описание особенностей формального выражения данных актантов (табл. 2 графа «*форма*»).

3. Схематическое описание основных структурных схем предложений, предикат которых представлен данным семантическим классом глаголов, то есть минимальных структурных схем (табл. 2 графа «*ММС*»).

Опишем подробнее приведённые выше шаги:

1. Определение валентности и валентного гнезда предиката.

Валентность предиката как способность глагола удерживать при себе определенное число актантов в предложении определяет количество правосторонних и левосторонних членов синтаксической конструкции. Носители субстанциональной семантики

(актанты), как уже упоминалось выше, могут занимать по отношению к предикату семь ячеек – одну левостороннюю и шесть правосторонних, семантика, форма и место которых строго определены предикатом.

Таблица 2 – Основные правила формальной реализации валентностных свойств глагольных предикатов (фрагмент приложения)

Семантические классы	Валентность	Форма	МСС	Пример
I. Действие и деятельность				
1.1. Движение субъекта				
1.1.1. Поступательное движение субъекта				
1.1.1.1. Предложения, отображающие ситуацию однонаправленного движения, ориентированного относительно исходного и конечного пунктов	1 5 6 1^0 5 6 1^0 5 6	1- N1 5 - N(из, с, от)2 6 - N(в)4; N(до)2 1^0 - N3; N0 5 - N(из, с, от)2 6 - N(в)4; N(до)2 1^0 - N0 5 - N(из, с, от)2 6 - N(в)4; N(до)2	N_1V_f Inf Vpl 3	Бежать Ехать Идти
1.1.1.2. Предложения, отображающие ситуацию однонаправленного движения, ориентированного относительно исходного пункта	1 5 1^0 5	1- N1 5 - N(из, от, с)2 1^0 - N3; N0 5 - N(из, от, с)2	N_1Vf Inf	Выходить Выезжать Отдаляться Отходить Удаляться Уходить
1.1.1.3. Предложения, отображающие ситуацию однонаправленного движения, ориентированного относительно конечного пункта	1 6 1^0 6	1 - N1 6 - N2; N(до)2, N(в)4 1^0 - N0 6 - N2; N(до)2, N(в)4	N_1Vf Inf	Достигать Добегать Вбегать Влезать

Кроме того, следует обратить внимание на то, что часто валентность глагола может реализовываться не полностью. Поэтому следует отличать обязательную и потенциальную валентность. Обязательная валентность является основой структурной схемы предложения и, соответственно, обязательные правосторонние компоненты предложения не могут быть опущены. Потенциальная же валентность в определенном контексте может быть нереализованной, но даже при этих условиях смысл будет понятен.

Для создания системы формализованных правил реализации предикатов в простых предложениях необходимо выполнить следующий алгоритм:

– спрогнозировать возможное количество ячеек предикатов определённых семантических классов, с учетом того, что максимальная валентность предиката – семь ячеек;

– каждая валентная ячейка предиката обозначается определенной цифрой (от *1* до *7*) и предусматривает свою конкретную формальную и семантическую реализацию в предложении;

– возле цифры, обозначающей ячейку, являющуюся обязательной для заполнения, ставим дополнительную метку.

Так, *1 2 3 4 5 6 7* – цифровые индексы, указывающие на необходимое заполнение определенных валентно обусловленных ячеек (*1* – левосторонний актант, или субъект действия; *2, 3, 4, 5, 6, 7* – правосторонние актанты и, соответственно: объект, адресат, инструмент, исходный, конечный, промежуточный локативы).

Например: предикат семантического класса «*отображающие ситуацию однонаправленного движения, ориентированного относительно исходного и конечного пунктов*» – «*бежать*» имеет валентность *1 5 6*

Пояснение: слово «*бежать*» является трехвалентным предикатом с заполненными первой, пятой и шестой ячейками.

В работе также была разработана система меток для обозначения дополнительных валентностных характеристик предиката:

– *1…. – звездочка при цифровом индексе указывает на необязательное заполнение данной валентно обусловленной ячейки предиката;

– 2_1, 2_2… – подстрочные цифровые символы указывают на возможное заполнение данной ячейки двумя актантами;

– 2(ся) – буквенный символ при цифровом символе 2 указывает на то, что заполнение данной ячейки валентности (с семантикой объекта) происходит в рамках слова и не требует дополнительного формального выражения;

– 1^0 – надстрочный символ «0» при цифре указывает на нулевое заполнение ячейки субъекта.

Например: предикат семантического класса «*отображающие ситуацию беспорядочного движения*» – *«метаться» имеет валентность* 1^0 **7*

Пояснение: 1^0 – нулевое заполнение ячейки субъекта

7-я ячейка предиката не обязательно должна быть заполнена для получения завершённого семантического целого.

или предикат семантического класса «*отображающие ситуацию избавления» – «избавлять*», имея валентность $1\ 2_1\ 2_2$ **5*, допускает вариативность при заполнении второй ячейки;

2. Описание особенностей формального выражения актантов.

Говоря о валентных связях, И. Выхованец отмечает, что «теория падежей и теория валентности – по сути две части той же теории» [10, с. 41], ведь заполнение валентных ячеек предикатов происходит при помощи актантов, строго определённых в падежном отношении. Вариативность же объектных форм в ячейках встречается довольно редко.

Для формального представления валентностных свойств предиката того или другого семантического класса в работе были приняты условные сокращения, отображенные в табл. 3.

Рассмотрим подробнее специфику заполнения ячеек при левосторонней и правосторонней валентности. Так, первая (левосторонняя) ячейка включает субъектную семантику, которая представляется именной частью речи в именительном падеже или, реже, инфинитивом, который является формальным субъектом.

Например: одновалентный предикат семантического класса «*отображающие ситуацию прекращения действия, бытия, состояния*» – «*прекратить*) реализуется по такому правилу: *1- N1*, то есть для данного предиката обязательно заполняемой является только одна левосторонняя ячейка субъекта.

Субъектный актант и предикат являются основой простого элементарного двусоставного предложения. *Например*: *Мир будет всегда*

Группа одновалентных глаголов является наиболее многочисленной. В большинстве случаев данные предикаты являются коррелятами к *2-, 3-, 4-, 5-, 6-* валентным глаголам.

Левосторонний актант часто является лексически невыраженным, что свидетельствует о нулевом заполнении ячейки субъекта.

Например: одновалентные предикаты «*греметь*», «*сквозить*», имея основную валентность *N1,* могут реализовываться по вариативному правилу *N0,* то есть левосторонний актант является прогнозируемым, однако формально невыраженным. Реализация в предложении: «*Гремит*», «*Сквозит*».

Лексическая невыраженность левостороннего актанта является вариантом реализации активного предиката:

«*он гремит*» – реализация активного предиката,

«*гремит*» – вариативная форма.

Далее рассмотрим валентно обусловленные актанты, несущие объектную семантику, т.е. актанты, занимающие правостороннюю позицию по отношению к предикату. Эти актанты чаще всего характеризуются признаком синтаксической необходимости, что касается в большей степени первых трех правосторонних ячеек, предусмотренных семантикой определенного глагола.

Например: трехвалентный глагол «*ругать*» имеет валентность *1 2 3 или* 1^0 *2 3, то есть 2 и 3 ячейки* обязательно должны получить реализацию в предложении.

Таблица 3 – Условные сокращения, принятые в работе для обозначения грамматических характеристик актантов

Буквенно-цифровое обозначение	Пояснение
N	имя существительное
N1 N2 N3 N4 N5 N6)	цифровой индекс при *N* обозначает номер падежа в парадигме по порядку
Adj	имя прилагательное
Adj1 Adj2 Adj3 Adj4 Adj5 Adj6	цифровой индекс при *Adj* обозначает номер падежа в парадигме по порядку
Adv	наречие
V_f	спрягаемая форма глагола
$V_{pl\ 3}$	форма множественного числа третьего лица глагола
$V_{s\ 3/n}$	форма единственного числа и 3 лица либо среднего рода глагола (в зависимости от формы времени
Inf	неопределенная форма глагола (инфинитив)
(при), (над, (из), (из-за), (в), (через)…	буквенные символы между *N* и цифровым индексом, обозначающим падеж имени существительного, называет предлог, с которым возможно заполнение данной ячейки

На семантико-синтаксическом уровне необходимо различать синтаксические категории субъекта, объекта, инструмента, локатива. Сигнификатом синтаксических категорий является обобщенное представление о типах ситуаций [9; с. 301].

Основанные на сочетаемости семантико-грамматические разряды слов позволяют выделить ряд категорий предикатов [9, с. 259-276]. Обратившись к принципам распределения глаголов по типам валентности, предложенных А. П. Загнитко в работе «Теоретична граматика української мови. Морфологія», произведем количественное и качественное структурирование предикатов по валентным гнездам.

Верхнюю позицию в иерархии категории валентности занимают шестивалентные и семивалентные глаголы, которые охватывают в своем потенциале всю падежную систему существительного. Речь идет о лексемах движения, которые предусматривают возможное заполнение правосторонних ячеек актантами с семантикой объекта, адресата, инструмента, исходного локатива, конечного локатива. Такие предложения включают в себя компоненты с обстоятельственной семантикой, которые выражаются наречием, инфинитивом, предложными формами и в типичных случаях находятся в препозиции.

Например: семивалентный предикат *семантического класса «отображающие ситуацию однонаправленного перемещения, ориентированного относительно промежуточного пункта» – «переправлять» имеет валентность*

*1 2 *3 *4 5 6 7* (вариант – 1^0 *2 *3 *4 5 6 7*)

Формально она реализуется следующим образом:

1- N1 (N0)

2- N4; N2

**3- N3*

**4- N5*

5- N(из, из-за)2; N(от)2; Adv

6 - N(в, за)4; N(до)2; Adv

7 - N(через)4

Наименее многочисленными являются группы пяти- и четырёхвалентных предикатов. Пятивалентные глаголы также характеризуются семантикой движения, за исключением того, что семантика таких глаголов не предусматривает актанта с семантикой инструмента (например, предикат «*носить*»). То есть в данном случае не реализуется творительный падеж.

Четырьмя актантами характеризуются глаголы, семантика которых прогнозирует «модификацию объекта». Здесь присутствуют актанты с семантикой инструмента, объекта с предлогом и объекта без предлога. Эти члены предложения выражают объектные семантико-синтаксические отношения. В типичных случаях представлены существительным.

Например: четырёхвалентный предикат семантического класса *«отображающие ситуацию собственно влияния» – «толкать»* имеет валентность *1 2 5 6 (*1^0 *2 5 6).*

Формально она реализуется следующим образом:

1- N1 (N3; N0)

2- N4; N2

5 - N(из)2; N(от)2

6 - N(в)4; N(до)2

Трехвалентные глаголы предусматривают заполнение ячеек актантами с семантикой объекта и адресата. То есть грамматически и семантически реализованными являются винительный и творительный падежи.

Например: предикат семантического класса «*отображающие ситуацию однонаправленного перемещения, ориентированного относительно исходного пункта*» – «*выводить*» имеет валентностную реализацию *1 2 5 6 (*1^0 *2 5 6)*

Формально он реализуется следующим образом:

1- N1 (N3; N0)

2- N4; N2

5 - N(из)2; N(от)2

6 - N(в)4; N(до)2

К двухвалентным словам относятся глаголы с семантикой бытия, создания объекта, отношения с ярко выраженной положительной или отрицательной оценкой. Кроме левостороннего актанта в данном случае обязательным является актант с семантикой прямого объекта в форме винительного падежа без предлога или в форме родительного падежа с предлогом.

Например: двухвалентный предикат семантического класса *«отображающие ситуацию эмоционально-оценочного отношения»* – *«любить»* имеет валентную реализацию: *1 2 (1^0 2),* формальная реализация выглядит следующим образом:

1- N1 (N3; N0)

2- N4, N(на)4, N3, N(к)3, N5, N(в)6

3. МСС – минимальные структурные схемы предложений.

Предикат определяет семантико-синтаксическую структуру анализируемого предложения. Именно выделение предиката в определённой синтагме становится первым шагом на пути корректного распознавания текста машиной. Выделенный предикат с указаниями возможного количества и качества ячеек-актантов будет являться потенциальной структурной схемой предложения. Наша задача в данном случае состоит в выделении минимального количества основных структурных схем предложения (основанном на описанном выше семантико-морфологическом анализе предиката), которые при машинном распознавании текста будут налагаться на реальные анализируемые предложения и таким образом проверяться на заполняемость ячеек.

Например: для предложения, содержащего двухвалентный предикат «любить», относящийся к семантическому классу «*отображающие ситуацию эмоционально-оценочного отношения*» и выраженный *Vf* (спрягаемой формой глагола), и левосторонний актант, выраженный *N1*, минимальная структурная схема будет иметь вид*: N1Vf.*

Одним из условий осуществления такого наложения является разграничение собственно актантов и «слов-обстоятельств» [9, с. 123-130]. Отделить факультативные элементы можно посредством отсечения предложных групп (как предложных дополнений, так и обстоятельств). Но обратим внимание на тот факт, что некоторые предложные группы являются наполнителями правосторонних ячеек и валентно связаны с предикатом. Поэтому они предварительно должны быть включены в адекватную модель валентности и тем самым не характеризоваться как факультативные.

Например: двухвалентный предикат семантического класса «*отображающие ситуацию внешнего проявления отношения*» – «*аплодировать*» (валентность *1 2*) имеет следующую формальную реализацию

1- N1

2- N4, N(на)4, N3, N(к)3, N(над, с)5,

где предлоги *на, к, над, с* являются частью предложных групп, валентно связанных с предикатом.

Проанализировав специфику создания валентных гнезд предикатов определённых семантических классов, а также формальную их реализацию, можно выделить основные структурные схемы предложений, создаваемых говорящим в процессе речи:

N1Vf – реализуется при существовании в предложении лексически выраженного правостороннего субъектного актанта. Субъектный актант и предикат являются основой простого элементарного двусоставного предложения. Семантически элементарными простыми предложениями называем предложения, которые состоят из одного предиката и определенных его семантико-синтаксической валентностью синтаксем существительных [10, с. 126].

Например: *Он поддерживал нас*.

Inf – реализуется при отсутствии в предложении лексически выраженного правостороннего субъектного актанта.

Например: *Не греметь.*

Vpl3 – реализуется при наличии предиката, выраженного формой множественного числа третьего лица глагола.

Например: *Строят и строят.*

$\mathbf{V_{s\ 3/n}}$ – реализуется при наличии предиката, выраженного формой 3 лица единственного числа или глагола среднего рода (в зависимости от формы времени):

Например: *Светает.*

Стоит также отметить, что каждый предикат может прогнозировать три минимальные структурные схемы предложения. Прогнозируемые же предикатом правосторонние актанты несут дополнительную семантическую нагрузку.

Таким образом, выделяя предикат в сегментах обрабатываемого текста, мы получаем минимальный набор прогнозируемых характеристик (семантических и грамматических), необходимых для корректного машинного понимания реального, естественно-языкового текста.

Выводы

В предлагаемой работе был разработан и описан валентностный подход к обработке естественно-языковых текстов, основывающийся на анализе предикативных текстовых структур. Для чего были разработаны и описаны основные правила валентностной реализации глагольных предикатов определённых семантических классов в простых предложениях русского языка, а именно: общие правила прогнозирования количества и качества актантов, принципы их формальной реализации, а также система минимальных структурных схем предложений.

Предложенный подход дает возможность значительно оптимизировать работу по машинной обработке естественно-языковых текстов, в частности, с помощью построения обобщенной модели предложения, на которую ориентируется говорящий, свести бесконечное множество создаваемых в процессе говорения текстов к некоторому набору минимальных структурно оформленных единиц, обладающих определенной семантической информацией, передаваемой говорящим.

Литература

1. Arion – информационно-аналитическая система [Электронный ресурс]. – Режим доступа : http://asknet.ru/Analytics/arion.htm
2. TextAnalyst 2.0 – персональная система автоматического анализа текста [Электронный ресурс]. – Режим доступа : http://www.analyst.ru/index.php?lang=eng&dir=content/products/&id=ta
3. АОТ – Автоматическая Обработка Текста [Электронный ресурс]. – Режим доступа : http://www.aot.ru/docs/synan.html#6-30
4. Теньер Л. Основы структурного синтаксиса / Теньер Л. – М. : Прогресс, 1988. – 656 с.
5. Москальская, О. И. Проблемы системного описания синтаксиса / О. И. Москальская. – М. : Высш. шк., 1981. – 175 с
6. .Кацнельсон С. Д. К понятию типов валентности / С. Д. Кацнельсон // Вопросы языкознания. – 1987. – № 3. – С.20-32.
7. Степанова М.Д. Части речи и проблема валентности в современном немецком языке/ Степанова М.Д. – М. : Высшая школа. – 1978. – С. 123-130.
8. Апресян, Ю.Д. Отечественная теоретическая семантика в конце XX столетия / Ю. Д. Апресян // Изв. РАН. – 1999. – № 4. – С. 39-53.
9. Загнітко А.П. Теоретична граматика української мови. Морфологія. / Загнітко А.П. – Донецьк : ДонДУ, 1996. – 300с.
10. Вихованець І.Р. Граматика української мови. Синтаксис / Вихованець І.Р. – К. : Либідь, 1993. –С. 123-124.

11. Есперсен О. Философия граматики / Есперсен О. ; [пер. с англ. В.В. Пассека и С.П. Сафроновой ; под ред. и с пред. проф. Б.А. Ильиша]. — М. : Издательство иностранной литературы, 1958.
12. Русские глагольные предложения: Экспериментальный синтаксический словарь / [под общ. ред. Л.Г. Бабенко]. – М. : Флинта: Наука, 2002. – 462 с.
13. Зализняк А.А. Грамматический словарь русского языка. Словоизменение / Зализняк А.А. – [изд. 5-е, испр.]. –М. : Аст-пресс, 2008.
14. Белошапкова В. А. Синтаксис // Современный русский язык / [под ред. В.А. Белошапковой]. – М. : Высш. шк., 1981. – С. 363-466.

Literatura

1. Arion – informacionno-analiticheskaja sistema [Jelektronnyj resurs]. Rezhim dostupa: http://asknet.ru/Analytics/arion.htm
2. TextAnalyst 2.0 – personal'naja sistema avtomaticheskogo analiza teksta [Jelektronnyj resurs]. Rezhim dostupa: http://www.analyst.ru/index.php?lang=eng&dir=content/products/&id=ta
3. AOT – Avtomaticheskaja Obrabotka Teksta [Jelektronnyj resurs]. Rezhim dostupa: http://www.aot.ru/docs/synan.html#6-30
4. Ten'er L. Osnovy strukturnogo sintaksisa. M.: Progress. 1988. 656 s.
5. Moskal'skaja O.I. Problemy sistemnogo opisanija sintaksisa. M.: Vyssh. shk. 1981. 175 s.
6. Kacnel'son S.D. Voprosy jazykoznanija. 1987. № 3. S. 20-32.
7. Stepanova M. D. Chasti rechi i problema valentnosti v sovremennom nemeckom jazyke. M.: Vysshaja shkola. 1978. S. 123-130.
8. Apresjan, Ju. D. Otechestvennaja teoreticheskaja semantika v konce XX stoletija. Izv. RAN. 1999. № 4. S. 39-53.
9. Zagnitko A. P. Teoretychna gramatyka ukrains'koi movy. Morfologija. Donec'k: DonDU. 1996. 300s.
10. Vihovanec' I. R. Gramatyka ukrains'koi movy. Syntaksys. K.: Libid'. 1993. S. 123-124.
11. Espersen O. Filosofija gramatiki. M. : Izdatel'stvo inostrannoj literatury. 1958.
12. Russkie glagol'nye predlozhenija: Jeksperimental'nyj sintaksicheskij slovar'. M.: Flinta: Nauka. 2002. 462 s.
13. Zaliznjak A. A. Grammaticheskij slovar' russkogo jazyka. Slovoizmenenie. M.: Ast-press. 2008.
14. Beloshapkova V.A. Sintaksis. Sovremennyj russkij jazyk. M.: Vyssh. shk. 1981. S. 363-466.

RESUME

E.A. Bondarenko, O.A. Kaplina

Principles of Automatic Processing of Natural Language Texts: Valentnostny Approach

In offered work valentnostny approach to processing of the natural language texts, based on the analysis of predikaktivkny text structures was developed and described. For what the basic rules of valentnostny realization of verbal predicates of certain semantic classes in simple sentences of Russian were developed and described, namely: general rules of a prognozikrovaniye of quantity and quality of actants, principles of their formal realization, and also system of the minimum block diagrams of offers.

The offered approach gives the chance to optimize considerably work on machining of natural language texts, in particular, by means of creation of the generalized model of the offer by which is guided telling, to reduce an infinite set of texts created in the course of a govoreniye to some set of the minimum structurally issued units possessing certain semantic information, transferred to the speaking.

Статья поступила в редакцию 03.01.2013.